Haptic Object Recognition in Underwater and Deep-sea Environments

Achint Aggarwal, Peter Kampmann, and Johannes Lemburg
DFKI GmbH Robotics Innovation Center (RIC), Bremen, Germany
e-mail: achintaggarwal.84@gmail.com, peter.kampmann@dfki.de, johannes.lemburg@dfki.de
Frank Kirchner
DFKI GmbH Robotics Innovation Center (RIC), Bremen, Germany and Robotics Group, University of Bremen, Bremen, Germany
e-mail: frank.kirchner@dfki.de

Received 31 May 2013; accepted 8 June 2014

Tactile sensors, because of their intrinsic insensitivity to lighting conditions and water turbidity, provide promising opportunities for perception and object recognition in underwater and deep-sea environments. However, the limited availability of tactile sensors for underwater use has led to limited research in this domain. Recently, we have developed a deep-sea-capable tactile sensing system, with high spatial and force resolutions, which has made underwater haptic exploration possible for the first time. This paper presents a tactile sensor-based object recognition and localization methodology for structured underwater and deep-sea applications. Our approach is based on database matching using a local feature-based Random Sampling and Consensus (RANSAC) algorithm, and sequentially evolving the resulting hypotheses over the course of object exploration. It can handle a large database of three-dimensional objects of complex shapes, and it performs a complete six degree of freedom localization of a static object. An approach to utilize both contact and free-space measurements is presented. Extensive experimentation is performed in underwater environments for validating both the sensor system and the algorithms. To our knowledge, this is the first instance of haptic object recognition and localization in underwater and deep-sea environments.

1. INTRODUCTION

Environment perception and object recognition in underwater and deep-sea environments is impeded by the inability of the standard off-the-shelf sensing systems, i.e., laser and vision sensors. While laser sensors and vision systems have proven their worth in many structured and even some unstructured ground-based applications, their utility underwater is inhibited by bad lighting conditions and water turbidity created either by dirt or by sediment stirred by the robot itself. Tactile sensors are strong contenders for augmenting these commonly used sensing modalities, because of their intrinsic insensitivity to lighting conditions and water turbidity. However, research in underwater tactile perception remains limited because of the scarcity of underwater tactile sensors.

Vision and range sensors have charge-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) arrays that yield one-shot environment scans. On the other hand, tactile data acquisition is a slow process since data are collected only via direct contact between the tactile sensor and its environment. This is the most important limitation of tactile sensors. Thus, it is desirable to have dense clusters of tactile sensing elements on robot appendages for maximum object data acquisition during systematic environment exploration.

This article focuses on the problem of tactile sensor-based object recognition and complete six degree of freedom (DOF) localization in structured underwater environments. We focus on the recognition of static objects, from a pre-known but large object database. We have developed a tactile sensing system with high spatial and force resolutions (Kampmann & Kirchner, 2012) that is capable of underwater use (Kampmann & Kirchner, 201x). Further, the usability of the system has been tested in deep-sea environments, up to ambient pressures of 600 bar (6 km underwater). We have mounted dense arrays of these tactile sensors on the appendages of a three-fingered robot gripper (called SeeGrip) as shown in Figure 1. The SeeGrip system is mounted on an industrial standard and deep-sea-capable Orion7P manipulator from Schilling Robotics (Figure 3). While the manipulator and gripper actuators enable the systematic exploration of new and unexplored areas of the object surface, their kinematics and angular encoders enable the fusion of these data collected over several contact iterations into an object map.

Direct correspondence to: Achint Aggarwal, e-mail: achint aggarwal.84@gmail.com

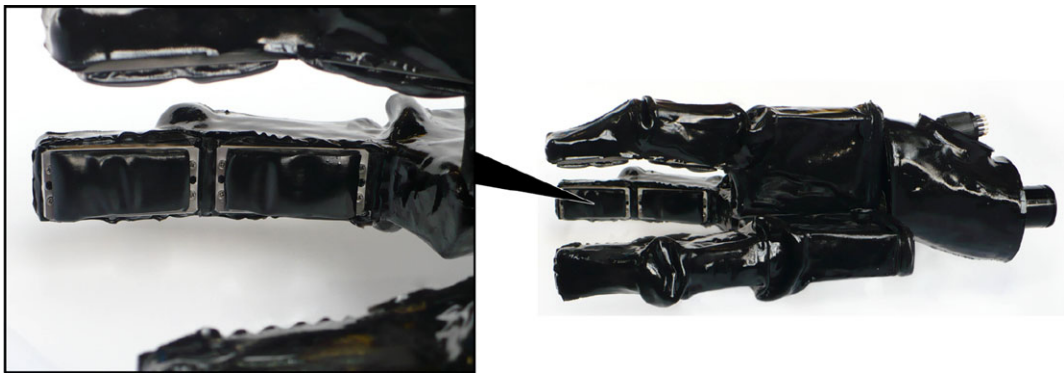

Figure 1. SeeGrip: A deep-sea-capable hydraulic gripper with tactile sensing units beneath a flexible skin covering.

The high force resolution of each sensing element of the SeeGrip sensors makes it possible to go beyond a binary contact or no contact detection, and it enables the estimation of the exact position of contact (Section 3.2). Further, the high spatial density of these sensing elements allows the contact positions to be combined together as point clouds. These locally dense point clouds can be used to estimate the local shape features of an object's surface. Thus, the feature-matching-based point cloud registration methods (Mian, Bennamoun, & Owens, 2005), common in range sensing applications, can be used for haptic object recognition. The difference, however, is that the tactile data are only sparsely spread around the object's surface. Thus, the data collected in the initial few exploration steps are generally insufficient for recognizing the correct object. Also, the low end-effector (EEF) position precision of industrial underwater manipulators leads to inaccuracies in the object map when data collected from multiple measurements are fused together.

Point cloud registration methods such as Random Sampling and Consensus (RANSAC) (Fischler & Bolles, 1981) can be used for matching the cumulative input point cloud with the database object point clouds for estimating the object's identity and pose. This procedure is generally nonsequential, and the knowledge or confidence in the object state achieved via point cloud registration at one exploration step does not influence the result of point cloud registration at a subsequent exploration step. On the other hand, since object exploration is a relatively slow process, the time during which the robot moves can be used for sequentially tracking and evolving the hypotheses.

We have therefore developed a heuristic approach for robust object recognition and localization. Local feature-based database matching, after every data collection step, generates a set of probable hypotheses. These hypotheses are evolved sequentially using new tactile measurements. For every hypothesis, the iterative closest point (ICP) (Besl & McKay, 1992) between the current cumulative measurements and the corresponding database object point cloud is used for finely correcting the hypothesis pose, thereby utilizing the additional measurement data. All hypotheses are then assigned relative weights by comparing the expected measurements for each hypothesis (using the object database) with the actual tactile sensor measurements. We use both the contact information and free-space information (from the swept volume of the robot) for this evaluation. At each exploration step, a fraction of the lowest ranked hypotheses are replaced by new hypotheses generated from feature-based database matching.

The feature-based database matching component allows the hypotheses to concentrate only on the high probability regions of the seven-dimensional space [one dimension for object recognition and six dimensions for localization of three-dimensional (3D) objects]. Sequentially evolving the hypotheses provides robustness and leads to a convergence to the correct hypothesis within a few measurements. Using ICP ensures that if an object pose is detected in the vicinity of the correct pose, it will eventually be corrected to the actual pose while evolving with new measurements. This also ensures that if a correct hypothesis was sampled at any time step, it will not be lost after subsequent measurements, which cannot be ensured using non-sequential point cloud registration alone.

In this article, approaches for object recognition and 6 DOF localization for structured deep-sea and underwater applications are presented for the first time in the literature. An overview of the robot system, including the tactile sensing principle, is presented first, followed by a discussion of the recognition methodology. The objective is to recognize and localize an object as quickly as possible using minimal tactile data. We focus on pre-known, non-deformable objects, which are assumed to be fixed in the environment. We address the recognition of 3D objects of complicated shapes and their localization in full 6 DOF space. We also assume a relatively large database of 45 objects. For practical applications, recognition of pre-known, objects from a large database is a reasonable constraint. Further, since it is not practical to create a database from ground truth collected via underwater measurements, we ensure that our

methodology can deal with a database constructed from laser sensors in the air. This database can be constructed autonomously in simulation and is easily expandable. Finally, the algorithms and the sensor system capabilities are validated by extensive experimentation under water and in deep-sea environments modeled via a pressure chamber, which generates ambient pressure up to 600 bar. While the SeeGrip system has been introduced in part in previous publications, this article presents the first instance of experiments with the complete system in underwater environments.

This article is structured as follows. Related work is discussed in Section 2. The gripper and tactile sensing system are discussed in Section 3. The object exploration methodology is discussed in Section 4, followed by a discussion of the recognition and localization algorithms. Experimental data collection and results are presented in Section 5, followed by a final conclusion in Section 6.

2. LITERATURE REVIEW

Research in tactile sensor-based object recognition for underwater applications is rare. However, there exists a good prior art for tactile sensing for ground-based applications. Some researchers have concentrated on the reconstruction of the geometric shape of previously unknown objects using tactile sensors (Allen & Roberts, 1989; Bierbaum, Gubarev, & Dillmann, 2008a). This involves the exploration of an unknown object's surface using an exploration strategy, and definition of the object shape using geometric models such as superquadrics.

A few researchers have focused on the recognition of pre-known objects by representing the explored object as volumetric models and matching with a known object database (Caselli, Magnanini, & Zanichelli, 1994). Other researchers (Gorges, Fritz, & W, 2010a) use point clouds to represent tactile data and use ICP for object matching. These approaches have been shown to work with small object databases, and their efficiency decreases linearly with the number of objects in the database.

Machine learning approaches have also been used for haptic object recognition. Neural networks (Johnsson & Balkenius, 2007), Bag of Features (Schneider et al., 2009), and Self-Organizing Maps (Gorges, Navarro, Goger, & Worn, 2010b; Heidemann & Schopfer, 2004) have been used to identify objects in the haptic space (using finger geometry and tactile data) itself without building a 3D representation of the object. These approaches have been proven to work well for object recognition from pre-known object databases, and they are particularly suitable for situations in which tactile data from actual objects are readily available for creating a training data set.

Some methods for active exploration of objects have been proposed (Moll & Erdmann, 2003). Objects are actively grasped and rolled without slipping between fingers to infer object motion, and thus fuse the tactile data gathered from various contact positions. This approach is shown to work on convex objects with smooth surfaces and known friction properties, and for practical applications it requires advanced hardware and controllers that are capable of such operations.

Bayesian techniques for tactile sensor-based localization have also been previously reported. One recent approach (Pezzementi, Reyda, & Hager, 2011) presents the application of particle filters and histogram filters for 3 DOF localization of 2D objects using a dense array of tactile sensors, similar to the sensors on the SeeGrip hand. Occupancy maps of objects built with the real sensor data are used as a database. These occupancy maps are queried with tactile sensor measurements to update the probabilities of the beliefs. This approach is shown to work well with a database of five 2D objects that are assumed to be static. Another approach (Petrovskaya & Khatib, 2011) presents Bayesian methods for fast localization of a known object in complete 6 DOF space using just a single tactile probe. The method uses annealing with a particle filter and gradually scales precision from coarse to fine levels. This approach is also proven to be effective for the localization of a non-static, movable object in 6 DOF space.

A few researchers have discussed the exploration strategies required for object exploration. A probabilistic task-based approach has been presented (Hsiao, Kaelbling, & Lozano-Pérez, 2010) for selection of the next best exploration direction, which allows the achievement of the desired grasps. A state estimator keeps track of the current state of an object, and an action selector selects the most appropriate action that maximizes the likelihood of success. The action domain consists of a set of action sequences, and a look-ahead procedure is used for selecting the most appropriate action. Some intelligent approximations are presented for efficiently managing the branching for this look-ahead procedure. This method has been validated by localizing a known 3D object in 3 DOF space, and it presents an efficient link between the object exploration and recognition modules. Another relevant work (Hebert, Howard, Hudson, Ma, & Burdick, 2013) presents an *Information Gain* approach for selecting the next best exploration direction. This strategy is based on minimizing the entropy of the belief distribution, and it is proven to work well for the localization of a known object. Other approaches (Roberts, 1990; Caselli, Magnanini, Zanichelli, & Caraffi, 1996) present single-finger tactile exploration strategies for recognizing polyhedral objects. Some others (Allen, 1990; Chen, Rink, & Zhang, 1996) have investigated the procedures for haptic exploration of object features. Recently, some approaches (Bierbaum, Rambow, Asfour, & Dillmann, 2008b; Mazzini, Kettler, Guerrero, & Dubowsky, 2011) have been proposed to direct exploration toward the maximum unexplored regions. Another strategy (Gorges et al., 2010a) is adaptable to a specific feature exploration

or to the exploration of previously unexplored regions. We have previously presented a widest unexplored cone-based exploration strategy (Aggarwal & Kampmann, 2012) that tries to utilize all the degrees of freedom of the manipulator and gripper systems, and it can deal with unknown objects of complicated shapes.

One article (Mian et al., 2005) provides an excellent review of the state of the art in object recognition techniques from range-sensing applications. RANSAC (Fischler & Bolles, 1981) is the most widely used technique for registration of point cloud data. The literature consists of methods for local surface feature-based point cloud matching (Besl & Jain, 1985; Hebert, Ikeuchi, & Delingette, 1995; Stein & Medioni, 1992), where the sensor data are rich enough to enable reliable local surface feature estimation. At the other end of the spectrum are global shape matching algorithms (Park, Germann, Breitenstein, & Pfister, 2010; Wahl, Hillenbrand, & Hirzinger, IEEE) that are relevant to applications in which sufficient information about the global shape of objects is available. Finally, there are methods (Drost, Ulrich, Navab, & Ilic, 2010) that combine local features and global shape estimation to recognize database objects. A local feature-based database pruning approach (Shan et al., 2004) allows efficient querying of large object databases.

3. SYSTEM DESCRIPTION

In this section, our three-fingered gripper (called SeeGrip) with multimodal sensing functionality is discussed briefly. We only focus on the parts relevant to our experiments on object exploration and recognition.

3.1. Gripper System and Electronics for Deep-Sea Operation

The SeeGrip (Figure 1) is a three-fingered deep-sea-capable hydraulic gripper that has been designed to work under ambient pressure of at least 600 bar, which corresponds to a depth of 6 km under water. Detailed information regarding the hydraulic finger actuation is provided in an earlier publication (Lemburg, Kampmann, & Kirchner, 2011). The gripper is covered by a fiber reinforced silicone skin with varying stiffness and thickness, depending on the region on the gripper surface. The gripper and the actuation principle have been tested under ambient pressure of up to 600 bar. For experimentation, this hydraulic system is integrated into the hydraulic circuit of the Orion7P manipulator to which the SeeGrip hand is attached (Figure 3), thus sharing the same fluid. The Orion7P hydraulics' supply pressure of 200 bar is reduced by a pressure control valve within the SeeGrip wrist to 50 bar working pressure.

As commercially available electronic components are not tested and rated for use in deep-sea environments by their manufacturers, experiments have been done in order to find the manufacturing techniques and materials that ideally withstand the ambient pressure. A hybrid approach has been used for realizing pressure tolerant electronics. First of all, miniature and inherently pressure-insensitive electronic components such as ceramic capacitors and microelectromechanical oscillators have been used. Secondly, the pressure-sensitive components on a printed circuit board are housed in a pressure-resistant hull. Thirdly, the electronic components are cast in epoxy resin, instead of polyurethane (PUR) or silicone, allowing additional structural support for single components that are inherently pressure-intolerant. To enhance the toughness of the otherwise brittle resin, powdered glass fibers are added and a slow and cold curing process is used. All these principles are proven with extensive testing of the SeeGrip system, with several cycles under ambient pressures up to 600 bar. Further details can be found in a previous publication (Kampmann, Lemburg, Hanff, & Kirchner, 2012).

3.2. Deep-sea Tactile Sensing System

Exploring objects in deep-sea environments poses the challenge of finding suitable measurement principles that can withstand the high ambient pressure of the water column. As the system is designed to be operated at up to 600 bar ambient pressure, a relative force measurement is preferred. A list of different types of sensors in the SeeGrip hand is given in Figure 2 and is shown graphically in Figure 4. Specialized sensors for static and dynamic forces as well as absolute force measurements have been selected. All sensors have proven their deep-sea capability. A fiber-optic measurement principle has been selected for detecting the geometric properties of the objects in contact, and it is the most relevant sensor for this article. Its working principle is therefore discussed below. More information on the other sensing modules can be found in other publications (Kampmann & Kirchner, 201x, 2012).

The working principle of the fiber-optic sensors is depicted in Figure 5. Each sensor element (sensel) consists of two optical fibers and a foam material. The endings of the optical fibers are placed within the foam or directly in front of it. One of the optical fibers leads light into the cell structure of the foam. The second optical fiber works as a brightness sensor. The scattered light in the foam structure is sensed and transferred to a phototransistor. Application of force to the foam structure leads to a compression of the foam, which results in a change in the scattering of the emitted light. This change is sensed and transferred to the phototransistor. The intensity of the sensed light is directly proportional to the amount of compression of the foam in front of the fiber-optic element. To create a small and simple setup, the phototransistor for each sensor element is realized by using a CMOS camera chip, without any optics. The fibers are directly glued to the CMOS chip die.

Sensor Type	Sampling Frequency	Sensor data	Number of Sensors	Output per second
Fiber optic sensor	30 Hz	640x480x8 Bit	12	210.94 MB/s
Piezoelectric sensor	10 kHz	20x16 Bit	2	781.25 kB/s
Force-torque sensor	1 kHz	6x16 Bit	1	11.72 kB/s
Absolute angular encoder	1 kHz	12 Bit	2	2.93 kB/s
Pressure sensor	1 kHz	14 Bit	4	6.84 kB/s
Temperature sensor	1 Hz	12 Bit	4	48 B/s
Total Data				211.72 MB/s

Figure 2. Sensor types used in the SeeGrip hand and the amount of sampled data without local preprocessing for one finger.

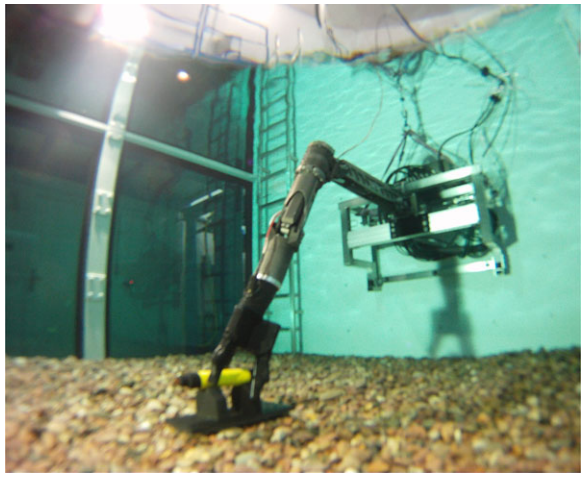

Figure 3. SeeGrip gripper on the Orion7P manipulator in the DFKI RIC Underwater Testbed.

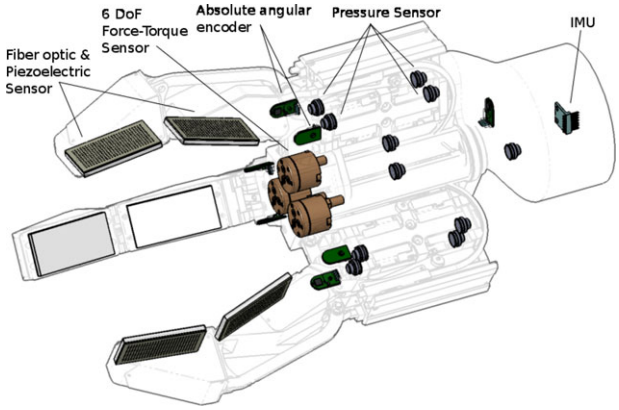

Figure 4. SeeGrip sensing system.

Sensor modules, each consisting of an array of 72 sensel elements, with a spatial resolution of 4 mm, have been developed. The sensors have a base size of 6 cm by 3 cm, with the foam thickness being 4 mm. Since these tactile sensors vary in their sensitivity from sensel to sensel, a calibration routine has been developed to map the amount of foam compression to the observed output for each sensel. Thus, each

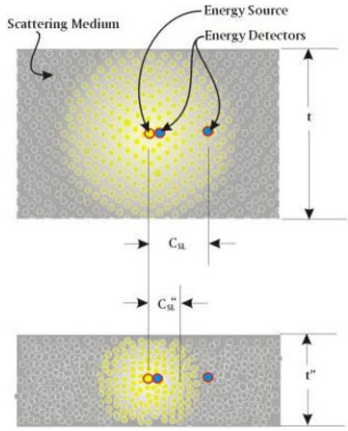

Figure 5. Working principle of the fiber-optic sensor (Reimer & Baldwin, 1999).

sensel is scaled individually and can output the amount of compression of the part of the foam directly in front of it. Finally, force-torque sensors at the base of each finger sense the absolute applied forces on each finger and are used to detect contact with the finger at a basic level.

For observing the signal response, a fiber-optic sensor with 72 sensel elements was tested on a Metaksa Testbench shown in Figure 6. The three-axis Testbench allows the precise positioning of a contact tip on the sensor unit such that each sensel can be compressed by the desired amount. A force sensor monitors the applied contact force. A 5 mm by 5 mm tip was used to gradually compress each individual sensel one by one. The response of the sensels was recorded along with the actual amount of foam compression and the applied force. The mean sensel response for all sensels (with standard deviation) is shown in Figure 7 with respect to the foam compression and the applied force. It can be observed that the sensor signal increases steadily with increasing foam compression. However, after about 80% of the foam compression (total thickness 4 mm), the sensor response drops. This is because the intensity of light reflected in the foam starts decreasing after the foam has been compressed beyond 80% of its thickness. This plot shows that the sensel signals can be directly used to interpret the amount of foam

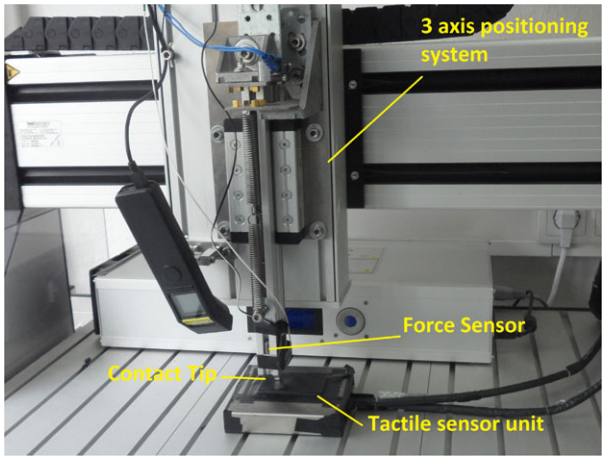

Figure 6. Metaksa Testbench for testing sensor characteristics.

compression. The corresponding sensor response versus the applied force is also shown for clarity.

For underwater and deep-sea operations, the sensor foam is filled with translucent silicone oil, which also fills the free spaces between the internal gripper components and its outer skin. Application of external force on the sensor squeezes the silicone oil out of the foam, and removal of the force leads to retraction of the foam and seepage of silicone oil back into the foam structure. The foam behaves like natural sponges that live at depths of up to 5,000 m underwater, and thus its natural structure is preserved. Additionally, the freely flowing silicone oil inside the skin also serves as an insulator for the electronics from possible short circuits, in case of water seepage inside the skin.

The necessity of using silicone oil within the sensor setup leads to some changes in the behavior of the measurement principle as compared to the air-filled foam. Since the sensor principle is based on the scattering of light in the foam structure, using silicone oil in the foam rather than air results in a dampening of the light feedback on the phototransistor. With a silicone filled foam with 4 mm thickness, compression changes below half a millimeter can be reliably sensed.

Sampling all the sensor information within the SeeGrip hand results in a lot of data that have to be transmitted to other processing systems. As shown in Figure 2, for one finger this would result in more than 200 MB per second of data. Multiplying this by the number of fingers on the gripper would result in more than 600 MB of data that need to be transmitted to a central processing unit every second. Therefore, a pre-processing strategy has been implemented that is responsible for the collection, interpretation, filtering, and fusion of the data from various sensors. All this is implemented on processors inside the gripper itself, and only filtered, high-level data are transmitted to an external computer. The hydraulic actuator control has also been implemented on the gripper processors, which manages the opening and closing of the fingers while using inputs from the desired sensors. Further information can be found in another publication (Kampmann & Kirchner, 201x).

4. OBJECT RECOGNITION AND 6 DOF LOCALIZATION

This section presents the object recognition and localization methodology. Our widest unexplored cone-based object exploration strategy has been presented in a previous publication (Aggarwal & Kampmann, 2012). The sensor model and data fusion methods are discussed in Section 4.1. An autonomous object database creation methodology is presented in Section 4.2. The local feature and global shape-based database matching methodology is discussed in Section 4.3. Finally, the sequential object recognition methodology is discussed in Section 4.4.

4.1. Sensor Model and Data Fusion

4.1.1. Sensor Model

Tactile data are transmitted from the SeeGrip gripper to an external computer in the form of preprocessed tactile images. A tactile image contains the signal value for each sensel of a fiber-optic sensor array. Additionally, it also

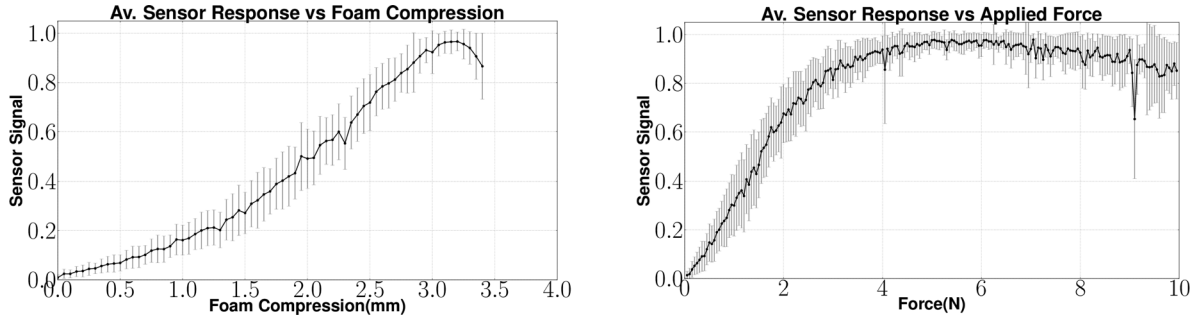

Figure 7. Mean sensor response (with standard deviation) of all sensels to foam compression (left) and applied force (right).

carries the absolute angular positions of all gripper joints at the time the tactile data were recorded. A sensor model is developed for estimating the point of contact in the local tactile sensor frame.

Let e_j denote the signal value for a sensel element $j \in [1, n]$, where n is the number of sensel elements in a sensor array. A threshold of 10% of the maximum expected signal $e_{\max}$ is used to filter out the noncontacting sensels. For the sensels on which contact is detected ($e_j > 0.1 * e_{\max}$), linear interpolation on e_j is used for estimating the compression of the foam, and subsequently the height of contact on the sensel z_j.

$$z_j = z_{\max} \frac{(e_{\max} - e_j)}{e_{\max}}. \tag{1}$$

Since the spatial position $(x_j, y_j)^L$ of sensel j with respect to the local sensor frame is preknown from computer-aided design (CAD) construction models, $(x_j, y_j, z_j)^L$ represents the complete position of a contact point in the local sensor frame.

4.1.2. Data Fusion: Contact and Free Space

The tactile data from all the sensors are transmitted to an external computer at the end of each data collection sequence. The sensor model discussed above is used to determine the contact positions $P_i^L = (x_i, y_i, z_i)^L,\ i \in [1, N]$ in the local sensor frames, where N is the total number of contact points. The forward kinematics of the gripper and manipulator is used to transform these local positions P_i^L to positions (P_i^W) in a fixed world coordinate frame using the following equation:

$$P_i^W = T_{\text{EEF}}(\Theta_{\text{arm}}) T_{\text{sensor}}(\Theta_{\text{finger}}) P_i^L. \tag{2}$$

In Eq. (2), $T_{\text{EEF}}(\Theta_{\text{arm}})$ is the homogeneous transformation defining the forward kinematics of the Orion7P arm up to its end-effector frame (EEF). $T_{\text{sensor}}(\Theta_{\text{finger}})$ is the homogeneous transformation of the sensor containing sensel i with respect to the EEF, and it defines the the forward kinematics of the corresponding finger.

All contact points $P_{i,t}^W$ collected at a time instance t are added together to form a point cloud $\text{PC}_t^{\text{cont}}$. To maintain a cumulative representation of the explored object's surface, a cumulative point cloud $\text{PC}_{\text{cum},t}^{\text{cont}}$ is maintained that consists of all measurements collected until time t. To add $\text{PC}_t^{\text{cont}}$ to $\text{PC}_{\text{cum},t-1}^{\text{cont}}$, all points in $\text{PC}_t^{\text{cont}}$ are simply appended to $\text{PC}_{\text{cum},t-1}^{\text{cont}}$, and a Voxel grid filter (Kaufman & Shimony, 1986) is used to filter out the overlapping or very closely spaced points.

For estimating the free space, a 3D occupancy map (Hornung, Wurm, Bennewitz, Stachniss, & Burgard, 2013) of the explored region is maintained. At each exploration step, this map is updated with contact point positions $\text{PC}_t^{\text{cont}}$. The swept volume of the gripper links is used as free space to update the occupancy map. For efficient update and search operations, the occupancy grid is maintained using an Octree representation.

At any time t, the free space can be represented as a free-space point cloud $\text{PC}_t^{\text{free}}$. This is done by querying the occupancy map for free-space Octree leaves within a bounding box B around the explored object. The size of the bounding box $\{B_x, B_y, B_z\}$ is defined by the maximum expected length of an object, $s_{\text{obj}_{\max}}$, along any dimension for the complete object database (see Section 4.2). At any time, a small bounding box B^{cont} with dimensions $\{B_x^{\text{cont}}, B_y^{\text{cont}}, B_z^{\text{cont}}\}$ is determined that completely contains all points in $\text{PC}_{\text{cum},t}^{\text{cont}}$. Then,

$$B_x = B_x^{\text{cont}} + 2s_{\text{obj}_{\max}}, \tag{3a}$$

$$B_y = B_y^{\text{cont}} + 2s_{\text{obj}_{\max}}, \tag{3b}$$

$$B_z = B_z^{\text{cont}} + 2s_{\text{obj}_{\max}}. \tag{3c}$$

Equation (3) ensures that for any possible object match hypothesis that contains at least one point from $\text{PC}_{\text{cum},t}^{\text{cont}}$, no surface contact point is possible outside the box B for any database object. The center of B is defined by the centroid of $PC_{\text{cum},t}^{\text{cont}}$. To generate $\text{PC}_t^{\text{free}}$, the free-space Octree leaves are sampled at a fixed resolution res_{free}, such that only one point lies inside a cube of size res_{free}.

4.2. Object Database

Since most underwater and deep-sea robotic applications require the recognition of *known* objects, having a preknowledge of the expected objects in the environment is advantageous. The database consists of objects in the form of their surface 3D position point clouds at two different resolutions: fine and coarse. Each point of the coarse point cloud also carries a feature description, which encodes the characteristics of the local shape of the surface at that position. To maintain generality, two necessary requirements have been ensured. First of all, the database can be created autonomously and new objects can be easily added. Secondly, it is not necessary to have the object database created by the exact ground truth collected via real tactile sensors.

The database is created in simulation and requires a CAD model of the object. A simulated laser scanner is used to completely scan the object surface. For autonomous exploration of the complete object, the widest cone exploration strategy (Aggarwal & Kampmann, 2012) is used to position the laser scanner to divert attention to maximum unexplored zones. For escaping local minima, random directions are used after a fixed number of iterations. The complete object surface contact cloud is computed by combining the contact point clouds generated by the laser scanner and filtering with a Voxel grid filter to eliminate overlapping points. Finally, the features (Section 4.3.1) of this contact cloud are calculated. The object database consists

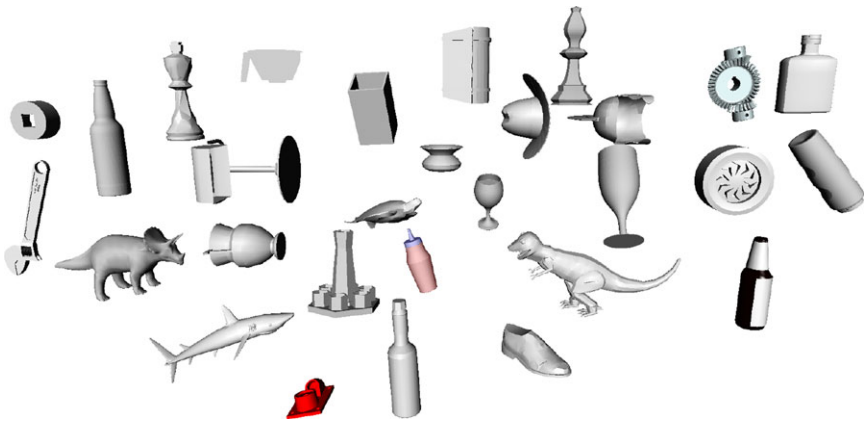

Figure 8. A few objects from the database of 45 objects.

of 45 objects (Figure 8), most of them from the Princeton Shape Benchmark Database (Shilane, Min, Kazhdan, & Funkhouser, 2004), with similar sizes (maximum length of 20 cm) such that they can be grasped by the SeeGrip gripper.

4.3. Local Features and Global Shape Matching

For fast and efficient database matching, a local feature and global shape-matching approach is employed. This algorithm combines feature-based model indexing for initial pruning of large databases [similar to Shan et al. (2004)] and geometric constraint based alignment (RANSAC) using only the pruned database for efficient object recognition.

4.3.1. Feature Selection

A variety of features and signatures have been presented in the literature that quantify the local or global shape of objects. For tactile sensing applications, it is necessary to choose features that are pose-invariant and have a high discriminative power that enables reliable point correspondence during database matching. Further, the feature computation should be quick to enable online computation of input cloud features. The descriptiveness of Spin Images (Johnson & Hebert, 1998), Signature of Histograms of Orientations (SHOT) (Tombari, Salti, & Di Stefano, 2010), Persistent Feature Histograms (PFH) (Rusu, Marton, Blodow, & Beetz, 2008), and Fast Point Feature Histograms (FPFH) (Rusu, Blodow, & Beetz, 2009) were compared for selecting the most appropriate feature. A database of nine different objects (two cubes of sides 5 and 10 cm; three cylinders of radii 3, 5, and 10 cm and height 5 cm; and four spheres of radii 3, 5, 10, and 15 cm) was prepared for comparison. These objects consist of 19 basic shapes representing the edges, vertices, and planar faces of the cubes; edges, curved surfaces, and planar surfaces of the cylinders; and surfaces of the spheres.

All four features of all relevant regions on the objects' surfaces were evaluated. The implementations of all four features in the Point Cloud Library (Rusu & Cousins, 2011) were chosen for comparison. The default parameters suggested by the authors were used. For Spin Images, the suggested parameters consist of eight bins along one dimension, 0° support angle, and 153 total dimensions for the feature histogram. For SHOT, the suggested parameters consist of 10 shape bins, 32 grid sectors, 28 angular sectors, 0 long-radius flow (lrf) radius, and 352 dimensions of the feature vector. For PFH, suggested parameters are 125 feature dimensions, and for FPFH, 33 feature dimensions were used. For normal estimation, a search radius of 1 cm was used, and for feature estimation, a neighborhood search radius of 1.5 cm was used, which corresponds to the maximum expected contact patch radius of the SeeGrip fiber optic sensor (3 cm width).

For comparison of the descriptiveness, for each feature the Euclidean distance between the feature vector values for all pair combinations of the 19 shapes was evaluated. These distances were plotted using heat maps as shown in Figure 9. The map for each feature varies from bright yellow to dull red, with the former representing a negligible difference and the latter representing a large estimate of distinctiveness. Thus, the capabilities of all four features in distinguishing between two given shapes can be judged by the color of the corresponding bin.

The SHOT features seem to detect only a negligible difference between curved cylindrical edges and straight cube edges (the corresponding bins are yellow). Also, the difference between large spheres and planar surfaces is large, which seems incorrect, as for the given parameters a sphere of 15 cm radius has a curvature comparable to a planar surface. The performance of FPFH and Spin Images seemed quite comparable except for a few cases. FPFH was able to distinguish between curved cylindrical surfaces of different radii much better than Spin Images. Similarly, FPFH can detect the difference between cylindrical edges of different

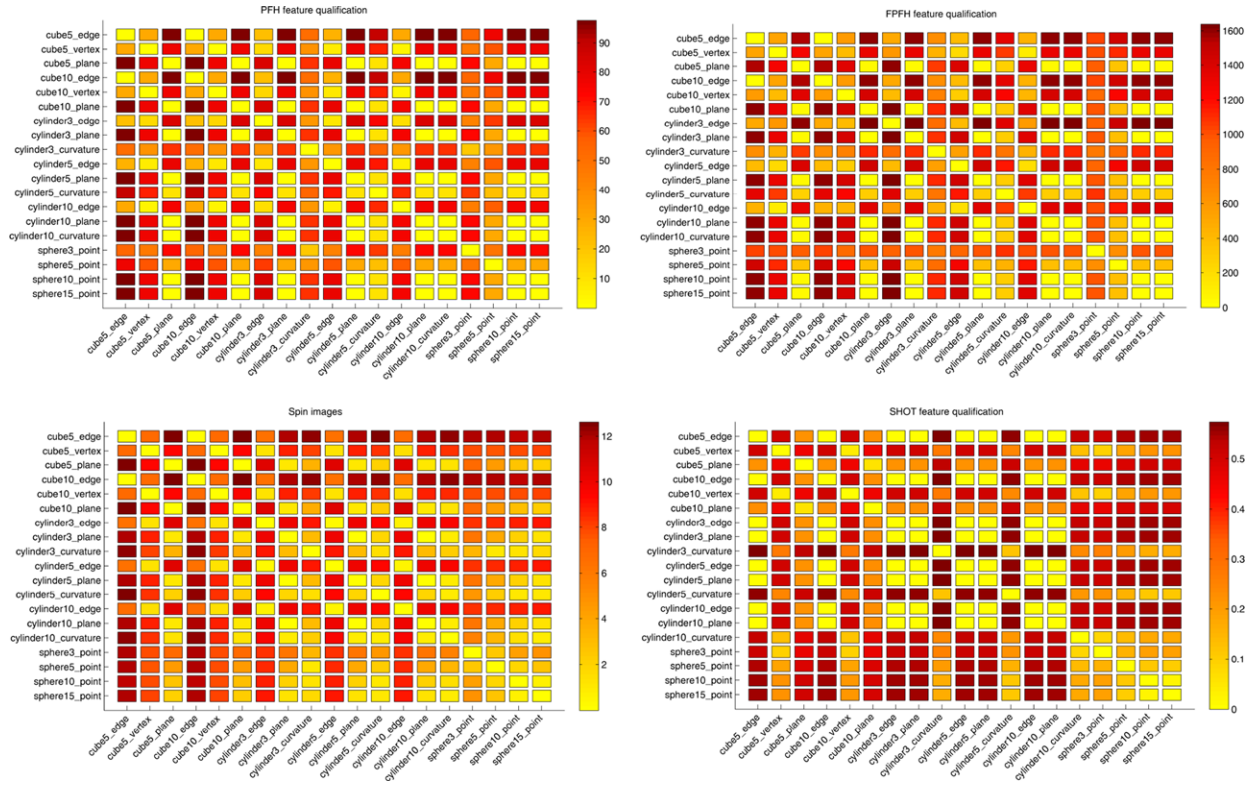

Figure 9. Discriminative characteristics of different features.

curvatures better than Spin Images. FPFH is also able to distinguish between spheres of different radii better than Spin Images. The performances of FPFH and PFH are very similar, with no critical differences. However, FPFH has a much shorter computational time and uses only 33 dimensions as compared to PFH with 125 dimensions and Spin Images with 153 dimensions. Thus, FPFH was chosen as the most appropriate feature for the tactile sensing application.

4.3.2. Batch RANSAC Algorithm for Database Matching

Initialization: For each database object "obj," the contact point cloud $\text{PC}^{\text{cont}}_{\text{db,obj}}$, its *k-d* tree (Muja & Lowe, 2009) $KT^{\text{cont}}_{\text{db,obj}}$, and the feature point cloud $\text{PC}^{feat}_{\text{db,obj}}$ are loaded. $\text{PC}^{\text{feat}}_{\text{db,obj}}$ for all objects are together loaded into a cumulative feature *k-d* tree $KT^{\text{feat}}_{\text{db,cum}}$. At any time t during object exploration, the exploration contact point cloud $\text{PC}^{\text{cont}}_{\text{cum},t}$ (Section 4.1.2) can be matched with the database.

Triplet sampling: Three points (p_1, p_2, p_3) are randomly sampled from $\text{PC}^{\text{cont}}_{\text{cum},t}$ such that they are spaced away from each other (using a predefined minimum distance threshold) and such that they do not lie on the boundaries of the point cloud patches (to facilitate reliable feature calculation). The features (f_1, f_2, f_3) for these three points are calculated.

Feature-based database pruning: For each feature f_1, f_2, and f_3, the closest matching W features are determined by querying $KT^{\text{feat}}_{\text{db,cum}}$. They are collected in three sets $F_{\text{db},1}$, $F_{\text{db},2}$, and $F_{\text{db},3}$, which together form the new reduced search space for registration.

Triplet registration: One feature point is randomly sampled from each of $F_{\text{db},1}$, $F_{\text{db},2}$, and $F_{\text{db},3}$, such that all three points belong to the same database object (say obj) and their corresponding points $p_{\text{obj},1}$, $p_{\text{obj},2}$, and $p_{\text{obj},3}$ in $\text{PC}^{\text{cont}}_{\text{db,obj}}$ have similar interpoint spatial distances as the input points p_1, p_2, and p_3. Singular value decomposition (Arun, Huang, & Blostein, 1987) between $(p_{\text{obj},1}, p_{\text{obj},2}, p_{\text{obj},3})$ and (p_1, p_2, p_3) is used to calculate the 6 DOF transformation $T = \{\text{pos, orient}\}$ between $\text{PC}^{\text{cont}}_{\text{db,obj}}$ and $\text{PC}^{\text{cont}}_{\text{cum},t}$.

Match evaluation: To evaluate the match, $\text{PC}^{\text{cont}}_{\text{cum},t}$ is transformed by T^{-1} to the coordinate frame of the database object (obj). This facilitates efficient error estimation using $KT^{\text{cont}}_{\text{db,obj}}$. For each point $c_i, i \in [1, C]$ in the transformed $\text{PC}^{\text{cont}}_{\text{cum},t}$, the distance d_{c_i} to the closest neighboring point in

$\text{PC}^{\text{cont}}_{\text{db,obj}}$ is evaluated using $KT^{\text{cont}}_{\text{db,obj}}$. The error contribution of each point c_i is given by

$$e^{\text{cont}}_{c_i} = \begin{cases} \left(\frac{d_{c_i}}{d_{\text{outlier}}}\right)^2 & \text{if } d_{c_i} < d_{\text{outlier}}, \\ 1 & \text{if } d_{c_i} \geq d_{\text{outlier}}, \end{cases} \quad (4)$$

where d_{outlier} is the outlier distance threshold.

The complete weight "wt" of the match is determined by

$$\text{wt} = \prod_{i=1}^{C} \exp\left(-\frac{1}{C} e^{\text{cont}}_{c_i}\right). \quad (5)$$

For each sampled triplet from the input, the *triplet registration* step is repeated R times to generate R object and pose hypotheses $S = \{s^{[0]}, s^{[1]}, \ldots, s^{[R]}\}$, where $s^{[i]} = \{\text{obj}^{[i]}, \text{pos}^{[i]}, \text{orient}^{[i]}, \text{wt}^{[i]}\}$, $i \in [1, R]$. $\text{pos}^{[i]}$ and $\text{orient}^{[i]}$ represent the 3D position and orientation, respectively.

Output: *Triplet sampling* and all subsequent steps are repeated N times to yield $N * R$ weighed hypotheses. The hypothesis with the maximum weight is considered as the best match. The corresponding obj is considered as the recognized object, and the 6 DOF pose {pos, orient} is considered as the recognized object's pose.

4.4. Sequential Object Recognition and Localization

In this section, a sequential object recognition approach is presented that is built upon the Batch RANSAC approach presented above, but it enables robust object recognition and pose estimation after only a few exploration steps. This is achieved by keeping track of the top ranked hypotheses from Batch RANSAC-based database matching at each exploration step and successively evolving them with new measurements from subsequent exploration steps. This is explained below.

Initialization: Using the initial contact point cloud, $\text{PC}^{\text{cont}}_{\text{cum},0}$, the Batch RANSAC algorithm (Section 4.3.2) generates an initial state $\mathcal{S}_0 = \{s_0^{[1]} \cdots s_0^{[n_p]}\}$, consisting of weighed object and pose hypotheses $s_t^{[i]} = \{\text{obj}^{[i]}, \text{pos}_t^{[i]}, \text{orient}_t^{[i]}, \text{wt}_t^{[i]}\}$, $i \in [1, n_p]$. The object index associated with a hypothesis does not change with time.

Hypotheses Evolution: The procedure for evolving the hypotheses $\mathcal{S}_{t-1}$ at time $t-1$, with new measurements to $\mathcal{S}_t$, is shown in Algorithm 1. The inputs for this algorithm are the set of hypotheses $\mathcal{S}_{t-1}$, cumulative contact point cloud $\text{PC}^{\text{cont}}_{\text{cum},t}$, cumulative free-space point cloud $\text{PC}^{\text{free}}_{\text{cum},t}$, and the database DB. In line 4, the pose of each old hypothesis is corrected using ICP between $\text{PC}^{\text{cont}}_{\text{cum},t}$ and the $\text{PC}^{\text{cont}}_{\text{db,obj}}$, the latter being transformed to the pose represented by the hypothesis, thereby utilizing the additional measurement data. Line 5 assigns an importance weight to the new hypothesis using the cumulative contact and free-space point clouds, which will be explained below. The newly formed weighed hypothesis is then added to an intermediate state $\overline{\mathcal{S}}_t$. In line 8, a new set of weighed hypotheses $\mathcal{S}_t^{'}$ is generated from the Batch RANSAC algorithm (Section 4.3.2). The *Evaluate* function similar to line 5, and as explained below, is used for the Batch RANSAC's match evaluation step in order to assign importance weights to the hypotheses of $\mathcal{S}_t^{'}$. In line 9, the top-ranked $\Psi * n_p$ hypotheses from $\overline{\mathcal{S}}_t$ are combined with top-ranked $(1-\Psi) * n_p$ hypotheses from $\mathcal{S}_t^{'}$ to form the final set $\mathcal{S}_t$.

Algorithm 1 Sequential hypotheses evolution algorithm

1: **procedure Evolve** ($\mathcal{S}_{t-1}$, $\text{PC}^{\text{cont}}_{\text{cum},t}$, $\text{PC}^{\text{free}}_{\text{cum},t}$, DB)
2: $\quad \mathcal{S}_t = \overline{\mathcal{S}}_t = \Phi$
3: $\quad$ **for** $k = 1$ to n_p **do**
4: $\quad\quad \{\text{pos}_t^{[k]}, \text{orient}_t^{[k]}\} =$ **ICP**($\text{obj}^{[k]}$, $\text{pos}_{t-1}^{[k]}$, $\text{orient}_{t-1}^{[k]}$, $\text{PC}^{\text{cont}}_{\text{cum},t}$, DB)
5: $\quad\quad \text{wt}_t^{[k]} =$ **Evaluate**($\text{obj}^{[k]}$, $\text{pos}_t^{[k]}$, $\text{orient}_t^{[k]}$, $\text{PC}^{\text{cont}}_{\text{cum},t}$, $\text{PC}^{\text{free}}_{\text{cum},t}$, DB)
6: $\quad\quad s_t^{[k]} = \{\text{obj}^{[k]}, \text{pos}_t^{[k]}, \text{orient}_t^{[k]}, \text{wt}_t^{[k]}\}$
7: $\quad\quad$ Add $s_t^{[k]}$ to $\overline{\mathcal{S}}_t$
8: $\quad \mathcal{S}_t^{'} =$ **BatchRansac**($\text{PC}^{\text{cont}}_{\text{cum},t}$, $\text{PC}^{\text{free}}_{\text{cum},t}$, DB)
9: $\quad \mathcal{S}_t =$ **Resample**($\overline{\mathcal{S}}_t$, $\mathcal{S}_t^{'}$)

At a given step t, the hypothesis $s_t^{[k]}$ in $\mathcal{S}_t$ with the maximum weight represents the recognized object obj_t^k and its 6 DOF pose $\{\text{pos}_t^{[k]}, \text{orient}_t^{[k]}\}$.

Hypothesis Evaluation: The purpose of the *Evaluate* function in Algorithm 1 is to assign an importance weight to a hypothesis. This is computed by using both the cumulative contact point cloud and cumulative free-space point cloud and is given by

$$\text{wt}\,(s_t, \text{DB}) = \text{wt}_{\text{cont}}\,(s_t, \text{DB}) * \text{wt}_{\text{free}}\,(s_t, \text{DB}). \quad (6)$$

In Eq. (6), $\text{wt}_{\text{cont}}(s_t, \text{DB})$ is the likelihood of the correctness of a hypothesis from contact measurements, and it is given by

$$\text{wt}_{\text{cont}}(s_t, \text{DB}) = \prod_{i=1}^{C} \exp\left(-\frac{1}{C} e^{\text{cont}}_{c_i}\right), \quad (7)$$

where $e^{\text{cont}}_{c_i}$ is the contact error associated with each point c_i, $i \in [1, C]$ in $\text{PC}^{\text{cont}}_t$ and is defined in Eq. (4).

For the free-space contribution, $\text{PC}^{\text{free}}_t$ (containing Q points) is transformed by the hypothesis pose. Then for each point $q_i, i \in [1, Q]$ in transformed $\text{PC}^{\text{free}}_t$, the distance d_{free_i} to the closest existing point in $\text{PC}^{\text{cont}}_{\text{db},obj}$ is calculated using $KT^{\text{cont}}_{\text{db,obj}}$. The free-space error contribution of each point q_i increases with reducing distance d_{free_i}, and is given by

$$e^{\text{free}}_{q_i} = \begin{cases} \frac{d^2_{\text{thresh}} - {d_{\text{free}_i}}^2}{d^2_{\text{thresh}}} & \text{if } d_{\text{free}_i} \leq d_{\text{thresh}}, \\ 0 & \text{if } d_{\text{free}_i} > d_{\text{thresh}}, \end{cases} \quad (8)$$

where d_{thresh} denotes the maximum distance threshold for d_{free_i}, beyond which the free-space point has no error contribution. This depends on the filter size used for estimating the free-space cloud (Section 4.1.2) and is given by $d_{\text{thresh}} = \frac{1}{2}\text{res}_{\text{free}}$.

Thus, the second term $\text{wt}_{\text{free}}(s_t, \text{DB})$ in Eq. (6) (the likelihood from free-space measurements) is given by

$$\text{wt}_{\text{free}}(s_t, \text{DB}) = \prod_{i=1}^{Q} \exp\left(-\lambda e_{q_i}^{\text{free}}\right), \tag{9}$$

where λ is a parameter used for adjusting the relative contributions of free space and contact space.

5. EXPERIMENTS AND RESULTS

Extensive experimentation was conducted in underwater and deep-sea (realized with a pressure chamber) environments for testing the gripper system as well as for validating the object recognition algorithms.

5.1. Underwater Tactile Data Collection

For experimental validation of the algorithms, five objects from the object database (Figure 8) were manufactured using a rapid prototyping machine. These objects are shown in Figure 10. The objects have been designed such that they can be fixed onto external fixtures using screws. For underwater experiments, the objects were rigidly mounted onto fixtures inside water. The SeeGrip gripper was attached at the end-effector of an Orion7P manipulator (Figure 3). Data were collected by manually steering the gripper to random positions on the object's surface and closing the grasp. The motion of the manipulator and the gripper takes around 1 min for a manually steered grasp. During the underwater experiments, only one fiber-optic tactile sensor at the distal link of the middle finger (shown in Figure 1) was available for generating contact information. The kinematics of all three fingers were used for generating free-space information. At least 50 random grasps were used to explore each of the five objects using this procedure. Some tactile images collected with the middle finger tactile sensor during these experiments are shown in Figure 19.

Figure 11 shows the complete tactile data collected for four objects using this procedure. The cumulative contact point clouds are shown at various data collection steps. This figure shows the sparseness of data collected by the tactile sensors (compared to laser scan data) even after 50 contacts with an object. The deformation of the shape of the cumulative contact point clouds gives a representation of the errors induced by imprecise EEF pose estimation of industrial underwater manipulators.

5.2. Description of Experiments

For evaluating the performance of algorithms, an object exploration run was modeled by using tactile measurements from random data collection grasps (Section 5.1) for each exploration step in the run. One exploration run consists of 40 exploration steps, and the results are averaged over several complete exploration runs for each object.

Object Recognition Rate: At each exploration step, the object recognition rate represents the percentage of times (over multiple complete exploration cycles) the correct object was identified as the top-ranked match.

Position and Orientation Errors: The position and orientation errors for each exploration step are also averaged over multiple exploration runs. The position estimation error is computed by evaluating the Euclidean distance between the actual and estimated point of origin of the object database point clouds. The orientation error is evaluated by estimating the absolute angle of rotation in 3D space between the actual object pose and the estimated object pose, using an axis-angle representation (Paul, 2008). Given a rotation matrix R between the two object poses, the angle of rotation θ can be determined using Eq. (10).

$$\theta = \arccos\left(\frac{\text{Tr}(R) - 1}{2}\right), \ 0 \leq \theta \leq \pi. \tag{10}$$

The position and orientation errors are computed only if the correct object is recognized at an exploration step. At every step, the position error is bound by a maximum value of 4 cm, and the orientation error by 1 rad, in case no correct object could be detected at this step for even a single exploration run.

5.3. Performance of Batch RANSAC

The Batch RANSAC-based object recognition and localization algorithm, discussed in Section 4.3, was evaluated using the tactile data collected from underwater experiments in Section 5.1 for the *Pitcher* object. A search radius of 7 mm was used for surface normal estimation and 10 mm was used for feature calculation. These dimensions have been chosen since the maximum possible radius of a contact patch generated by the sensor is 15 mm. The database consists of 45 objects. For database pruning $W = 3,000$ and for contact error estimation $d_{\text{outlier}} = 20$ cm were used.

Performance was evaluated for different combinations of parameters N(number of triplet samples from input) and R(number of samples in the pruned database for each input triplet). $N * R$ defines the total number of samples used in RANSAC. Figure 12 shows the results for (a) $N = 15$, $R = 200$ (3,000 samples); (b) $N = 50$, $R = 200$ (10,000 samples); (c) $N = 50$, $R = 500$ (25,000 samples); and (d) $N = 100$, $R = 1000$ (100,000 samples). The results are averaged over 50 complete exploration cycles, each consisting of 40 exploration steps represented by the horizontal axis. The plots show that the object recognition rate increases with an

Figure 10. Database objects printed using a rapid prototyping machine for hardware-based tests.

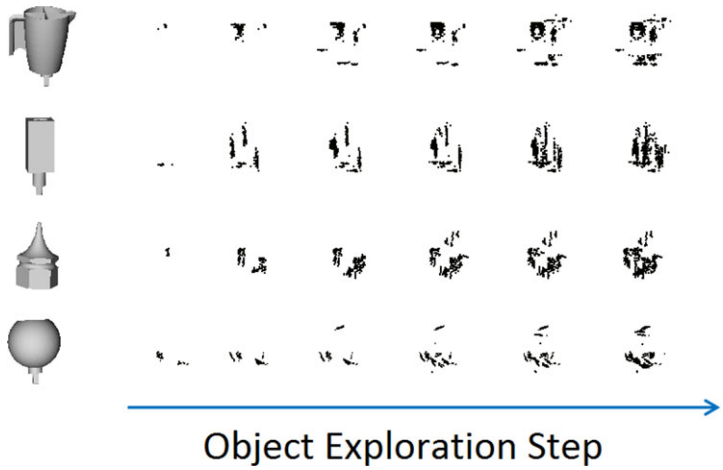

Figure 11. Cumulative data collected with increasing number of random grasps for all objects.

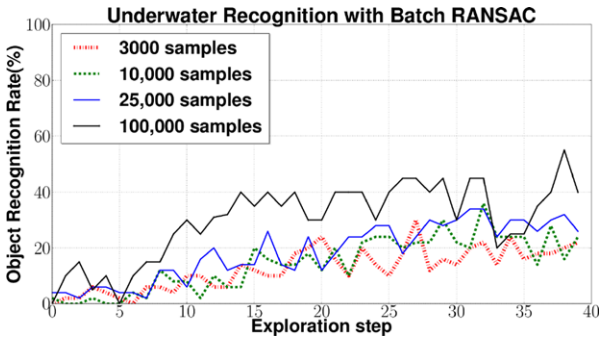

Figure 12. Object recognition performance of the Batch RANSAC algorithm for the object Pitcher with a 45-object database. Results are shown with varying number of samples for RANSAC and are averaged over 50 exploration runs.

increasing amount of the object's surface being explored. Also, as expected, increasing the number of samples increases the relative recognition rates. However, object recognition rates are unstable and only moderate even after a large part of the object has been explored, even for *case d* with 100,000 samples. The average computation times vary from 2 to 8 s for *case a*, depending on the number of points in the input point cloud, which increases steadily with increasing measurements. Average computation times were 8–25 s for *case b*, 12–40 s for *case c*, and 35–100 s for *case d*, all on a Quad-core, Intel I-7 processor with 8 GB RAM.

5.4. Augmenting Batch RANSAC with ICP and Free-space Information

To validate the utility of the sequential hypotheses evolution framework, it is necessary to evaluate the performance of the non-sequential Batch RANSAC algorithm when it is augmented with ICP as well as with free-space information. In the first set of experiments, ICP was used for augmenting the Batch RANSAC algorithm presented in Section 4.3. The pose of each hypothesis generated by the *triplet registration* step was finely corrected by using ICP between the

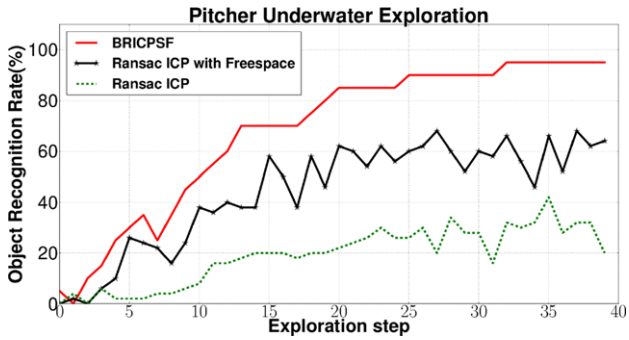

Figure 13. Comparison between object recognition rates for the object Pitcher with underwater exploration data when using BRICPSF (Case 1), ICP, and free-space augmented non-sequential Batch RANSAC (Case 2), and only ICP augmented non-sequential Batch RANSAC (Case 3).

current cumulative input contact point cloud $PC^{cont}_{cum,t}$ and the corresponding database object point cloud $PC^{cont}_{db,obj}$, after transforming the latter to the pose represented by the hypothesis. The hypothesis was then simply evaluated using the *match evaluation* procedure as was presented in Section 4.3. Similar to the previous cases, this methodology was tested using tactile data collected from exploring the object *Pitcher* under water, with a database of 45 objects. A total of 3,000 samples ($N = 15$, $R = 200$) were used for the RANSAC sampling phase. The object recognition results are shown with the dotted plot in Figure 13. The results were averaged over 50 complete exploration cycles, each consisting of 40 exploration steps. The average computation times vary from 3 to 30 s (between the 1st and 40th exploration steps).

In another experiment, along with the use of ICP for pose correction, free-space information was also used for evaluating the Batch RANSAC generated hypotheses. To incorporate the free-space information, after correcting a hypothesis pose with ICP (as was explained above), Eq. (6) was used for evaluating the hypothesis instead of Eq. (5) in Section 4.3. Using the same parameters as above, the results for this experiment are also shown in Figure 13 with a solid starred plot. The average computation times vary from 6 to 65 s (between the 1st and 40th exploration steps).

5.5. Performance of BRICPSF

For evaluating the BRICPSF algorithm, a number of hypotheses $n_p = 100$ and resampling parameter $\Psi = 80\%$ were used. For the free-space estimation $s_{obj_{max}} = 20$ cm, $res_{free} = 4$ cm, $d_{thresh} = 2$ cm, and $\lambda = 0.02$ were used. For its Batch RANSAC component, 3,000 samples ($N = 15$, $R = 200$) were used in the sampling phase.

Comparison with ICP and free-space augmented Batch RANSAC: Figure 13 shows the performance comparison between BRICPSF, ICP augmented Batch RANSAC, and ICP and free-space augmented Batch RANSAC for object recognition. Similar to the previous cases, the tactile data collected from exploring the object *Pitcher* under water were used with 45 database objects for BRICPSF evaluation. The results are averaged over 20 complete exploration cycles, each consisting of 40 exploration steps, represented by the horizontal axis. As was the case with the augmented Batch RANSAC, the BRICPSF algorithm also uses 3,000 samples for its Batch RANSAC component.

As can be seen in Figure 13, the object recognition rates are low in the beginning of exploration but increase gradually with a greater surface of the object being explored. The recognition rates using BRICPSF are much higher (about 50% better) and show a steady increase with increasing tactile measurements as compared to the relatively unstable behavior of the non-sequential Batch RANSAC. Around 90% success rates could be consistently achieved with BRICPSF, after only 20 exploration steps. The average computation times for BRICPSF vary from 19 to 80 s (between the 1st and 40th exploration steps). As before, results are generated on a Quad-core, Intel I-7 processor with 8 GB RAM.

Results for other objects: Object recognition, and position and orientation estimation results for the *Pitcher*, *Cuboid*, *Sphere*, and *Nut* objects are presented in Figures 14 (left) and 15, respectively. These results have also been generated using tactile data collected from the underwater experiments. For all four objects, the object recognition rates increase steadily, and the position estimation errors decrease steadily using the BRICPSF methodology. Object recognition rates of 80–100% were achieved after 15–20 exploration steps. Also, the position estimation error was estimated to be below 1 cm for all objects. The orientation error for *Pitcher* reduces steadily with increasing number of measurements and remains constant at approximately 5° after 20 exploration steps. For the other three objects, the plots show a large orientation error because of the symmetry in the shapes of these objects. Since the orientation error represents the absolute rotational error about the exact expected orientation of the object, it does not incorporate the effect of symmetry. However, the highest and lowest errors for the *Cuboid* and *Nut* objects can be seen to be consistently close to 0° and 180°. Since the *Sphere* is symmetric about multiple axes, the highest and lowest errors are randomly spread and illustrate that it is hard to estimate the exact orientation using shape information alone.

Changing database size: Next, the effect of changing the size of the database is evaluated. In Figures 14 (left) and 15, a database of 45 objects has been used to conform to a practical situation, where a large database is required. However, for comparison with other approaches in the literature, it makes sense to present results for a smaller database of 5 objects. The results are shown in Figures 14 (right) and 16. Reduction in the database size leads to a substantial improvement in the object recognition

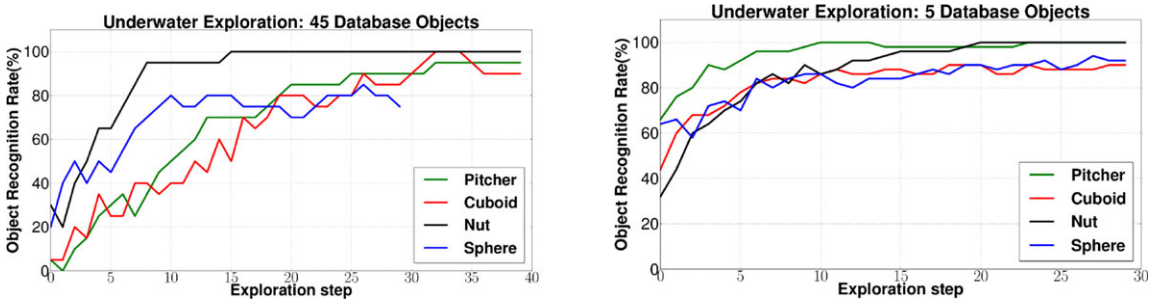

Figure 14. Underwater object recognition rates using BRICPSF for four different objects with a database consisting of 45 objects (left) and 5 objects (right). Results are averaged over 20 exploration runs.

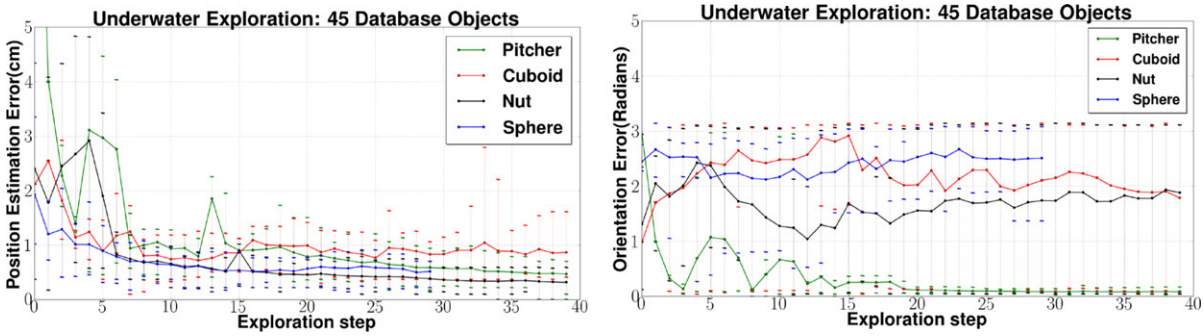

Figure 15. Underwater average position estimation error (left) and orientation estimation error (right) with 95% confidence intervals using BRICPSF for the objects Pitcher, Cuboid, Nut, and Sphere. The database consists of 45 objects, and results are averaged over 20 exploration runs.

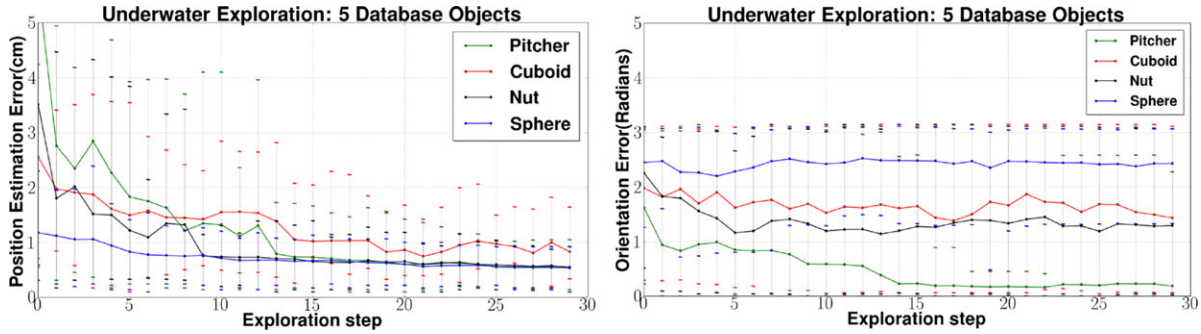

Figure 16. Underwater average position estimation error (left) and orientation estimation error (right) with 95% confidence intervals using BRICPSF for the objects Pitcher, Cuboid, Nut, and Sphere. The database consists of five objects, and results are averaged over 20 exploration runs.

performance. This is justified by the reduction in the search space. Object recognition rates of 80–100% were achieved after only 5 exploration steps for all the objects. The average position estimation errors were also estimated to be below 1 cm after 5–14 exploration steps for all the objects.

5.6. Deep-sea Experiments

The experiments under deep-sea ambient pressure conditions have been conducted in a pressure chamber at DFKI RIC. Figure 17 shows the pressure chamber and the experimental setup with the SeeGrip gripper and an object in the air. This pressure chamber has been used to generate ambient pressure of up to 600 bar. The usable space for experiments in the chamber has a diameter of 40 cm and a depth of 60 cm. A total of 48 electrical connectors provide an interface between the chamber and the external world. This allows the *in situ* operation of robotic systems. To actuate the SeeGrip system, a hydraulic supply has also been integrated into the pressure chamber. A

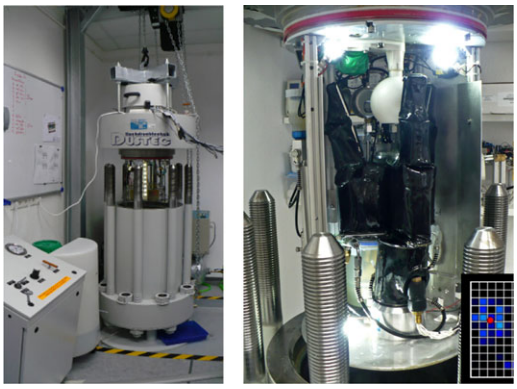

Figure 17. Experimental setup for the SeeGrip gripper in a pressure chamber. Tactile image recorded while touching the sphere object at 600 bar (right bottom).

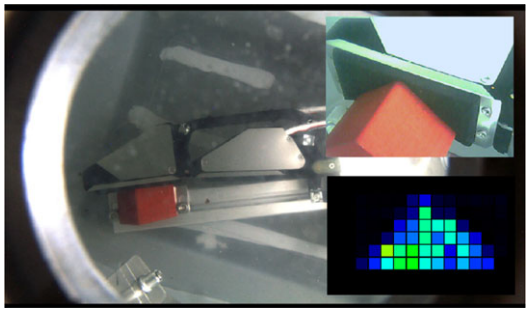

Figure 18. Feedback of the fiber-optic sensor at 600 bar while in contact with a triangular shape.

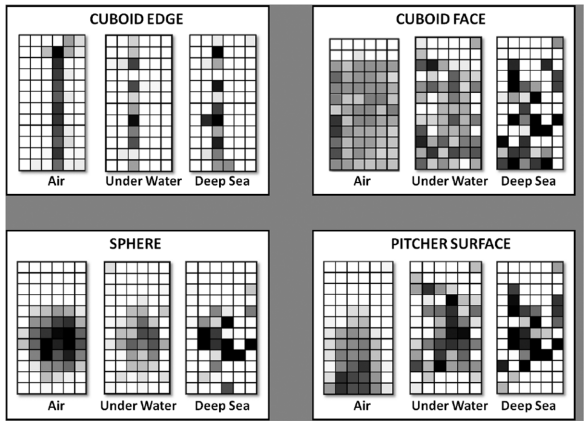

Figure 19. The difference between tactile images while exploring standard object surfaces for three different cases. Darker shades represent greater foam compression.

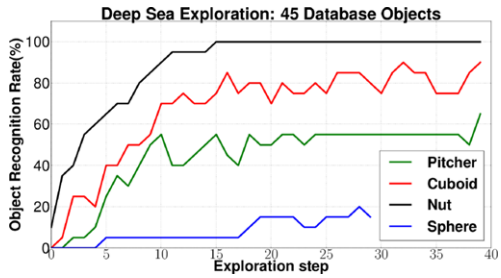

Figure 20. Deep-sea object recognition rates using BRICPSF for four different objects with a database consisting of 45 objects. Results are averaged over 20 exploration runs.

looking glass allows monitoring of the experiment from the top.

The SeeGrip gripper and fiber-optic tactile sensors have been proven to work repeatedly with the experiments in the pressure chamber up to ambient pressures of 600 bar. The tactile data collected at 600 bar by interactions between the fiber-optic sensor array and two different object surfaces are shown in Figures 17 and 18. These prove the effectiveness of the sensor in detecting contact under high ambient pressure.

Tactile Data: Since the area inside the pressure chamber is very limited for extensive object exploration, the data that were collected from the underwater experiments are used to evaluate the sensor performance in deep-sea. Since it has been proven that the tactile sensors can at least distinguish contact from no contact under high pressures, the spatial measurements of the tactile data that were collected in the underwater experiments can be safely assumed to be valid for deep-sea cases as well. The force resolution, however, is reduced by adding Gaussian noise of 50% to the force measurements of each contacted sensel. For a 50% force resolution noise, the covariance matrix to model the uncertainty as Gaussian white noise is chosen as $Q = \text{diag}(2\text{ mm}, 2\text{ mm}, 2\text{ mm})$, similar to Meier, Schöpfer, Haschke, & Ritter (2011). This results in tactile images of lower descriptiveness as shown in Figure 19. Since the Orion7P manipulator is a deep-sea-capable manipulator, the position errors stay the same as in the underwater experiments.

Results: The results of using the BRICPSF approach with 45 database objects and the same parameters as the underwater case are shown in Figures 20 and 21. The recognition rate reduces drastically for the *Sphere* object, and moderately for the other three objects. This is because the *Sphere* object consists of similar types of features throughout its surface, and reducing the descriptiveness of the tactile image results in very few matches in the feature-matching step. For the other three objects, 50–100% recognition rates were still achieved.

5.7. Discussion

5.7.1. Hardware performance

The extensive underwater experimentation with the SeeGrip fiber optic tactile sensors has validated their utility for underwater environment exploration. Further, the sensors

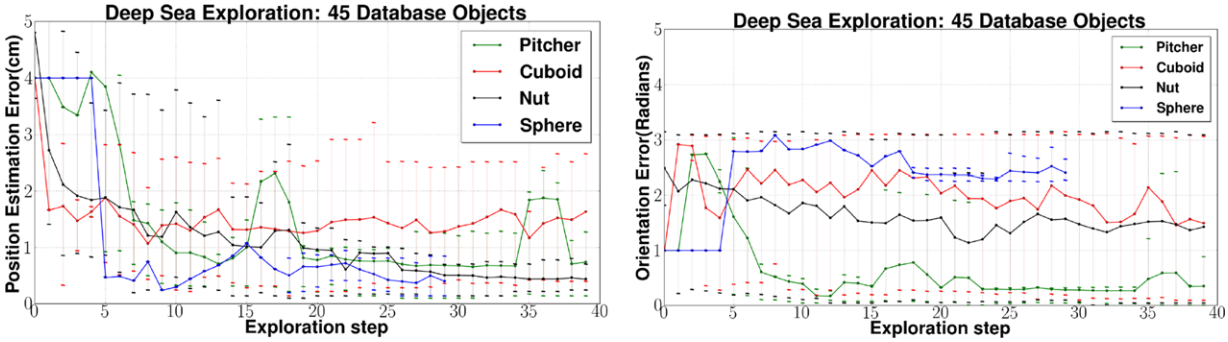

Figure 21. Deep-sea average position estimation error (left) and orientation estimation error (right) with 95% confidence intervals using BRICPSF for the objects Pitcher, Cuboid, Nut, and Sphere. The database consists of 45 objects, and results are averaged over 20 exploration runs.

have been proven to work in deep-sea-like conditions at ambient pressures up to 600 bar (corresponding to a depth of 6 km under water). The tactile measurements also demonstrate the effectiveness of the sensors in determining the precise point of contact within the 4 mm compression range of the foam (rather than just a binary contact or no contact detection). Using the forward kinematics of the gripper and manipulator, multiple measurements around an object's surface have been concatenated together to form a representation of the object (Figure 11). The reliability of the surface features estimation from contact point clouds is validated by the underwater object recognition performance. Further, the corruption of underwater tactile measurements with artificial noise (Section 5.6) resulted in degradation of feature quality, which eventually caused degraded object recognition performance (Figure 20 as compared to Figure 14). This confirms that reliable features were indeed estimated with the underwater tactile measurements.

5.7.2. Performance improvements with sequential hypotheses evolution

The experiments prove that sequentially evolving hypotheses using BRICPSF results in large object recognition and localization performance improvements over the stand-alone Batch RANSAC-based database matching methodology. The results in Figure 12 show that Batch RANSAC alone does not yield high recognition performance even after several measurements. Further, even increasing the computational effort by increasing the number of samples to 100,000 samples does not result in major improvements in the performance. This is because the success of this approach depends on the selection of the correct object matches, first during the feature-based database pruning step, and then on finding the correct triplet matches during registration. Thus, even if the database is pruned correctly, unless all possible object and 6 DOF pose hypotheses are tested comprehensively, there is a possibility of the correct object and pose not being detected.

The results for augmenting the non-sequential Batch RANSAC based database matching with ICP and free-space information were shown in Figure 13. While this results in an improved performance over stand-alone Batch RANSAC, the recognition rates are unstable and are still substantially worse than BRICPSF. This can be explained by the fact that since the Batch RANSAC algorithm does not involve sequential tracking of hypotheses, detecting the correct hypothesis at one instance during the course of exploration does not have any influence on the success at a subsequent exploration step. This leads to moderate and unstable object recognition rates even after a large part of the object has been explored.

BRICPSF, on the other hand, ensures that if, at any instance during the course of exploration, a hypothesis is detected in the vicinity of the correct object and the correct pose, it will be tracked and successively evolved to the correct pose. This results in high recognition rates (Figures 13 and 14). The use of feature-based RANSAC within the sequential framework allows an efficient handling of large databases. While the 45-object database consists of more than 90,000 object surface feature points, feature-based pruning allows the database to be pruned to only 3,000 points for each input sample, which represents only 3.33% of the database. With a large database of 45 objects, 80–100% recognition rates were achieved within 15–20 grasps for all four objects. Position estimation errors of less than 1 cm were also achieved. This is quite acceptable for underwater applications, given the inaccuracies in estimating the EEF position of industrial underwater manipulators (Aggarwal & Albiez, 2013).

5.7.3. Comparisons with other approaches

In this article, haptic underwater object recognition and localization has been addressed for the first time in the

literature. Comparisons can be drawn with other approaches for terrestrial environments. As mentioned in Section 2, the prior art consists of several interesting and effective approaches for haptic object recognition.

One relevant prior art (Pezzementi, Reyda, & Hager, 2011) is comprised of the application of particle filters and histogram filters for 3 DOF localization of 2D objects using a dense array of tactile sensors, similar to the SeeGrip sensors. Occupancy maps of objects built with real sensor-based ground truth are used as a database. The histogram filter with 10,000 bins was able to successfully achieve 100% recognition rates with nine sensor readings, using a database of five objects. The objects were localized to within 1.3 mm position error. Another approach (Petrovskaya & Khatib, 2011) presents 6 DOF localization of a known 3D object using a tactile probe. It uses an annealed particle filter and gradually scales precision to manage large dimensional spaces. This results in a localization accuracy of 5 mm (position) and 3° (rotation) in about 12 probing operations.

Another approach (Gorges et al., 2010b) achieves 80–100% recognition rates with a database of seven objects, using only five to six palpations. In this approach, however, recognition is performed directly in the haptic space by using the gripper kinematics and tactile sensor data, without building a representation of the object. Thus, object localization is not addressed. Similarly, another approach (Schneider et al., 2009) uses Bag of Features for object recognition and results in 85% recognition rates, with a database of 21 objects, in 10 grasps. Again, it does not deal with object localization. Also, both of these approaches depend on the ground truth data collected from real sensors for creating a training data set.

The results of our BRICPSF algorithm are also quite comparable. We were able to achieve 80–100% recognition rates with 5–10 grasps for a database of five 3D objects. Position estimation error was around 1 cm. These results are quite acceptable, since the position errors associated with an underwater manipulator and sensor system are much greater as compared to ground-based systems. Also, our approach does not depend on real ground truth data collected from the tactile sensors. However, the experiments conducted in this article were based on manually steering the manipulator and gripper systems for data collection. For autonomous exploration, an object exploration strategy should be used. For efficient exploration, this exploration strategy should be tied to the current object recognition state. The task driven exploration strategy presented in Hsiao et al. (2010) and information gain strategy (Schneider et al., 2009) are possible ways of providing an efficient link between exploration and recognition.

5.7.4. Autonomy in structured underwater environments

The problem of static object recognition for a structured underwater scenario has been tackled efficiently using an ICP augmented sequential hypotheses evolution methodology in combination with Batch RANSAC based database matching. It has been ensured that the algorithm can handle a large database, of at least 45 objects, and that the database does not depend on the actual ground truth tactile data. Both of these conditions are necessary for reasonably autonomous scenarios. A computational time of less than 1 min (up to the 20th exploration step) for each exploration step of the BRICPSF is also reasonable. This is quite comparable to the time taken by the manipulator to move to a new exploration position, and it can therefore be done in parallel to the robot motion. The experiments in this article have been performed while only one tactile sensor was operational. With multiple operational tactile sensors, even better performance can be achieved. The results presented in the current article can be considered as a step toward achieving autonomy in underwater and deep-sea environments.

6. CONCLUSION AND FUTURE WORK

This article presents a system-based approach for recognition and 6 DOF localization of pre-known objects using tactile data, for structured deep-sea and underwater applications. The design principles behind the construction of a deep-sea-capable three-fingered hydraulic gripper and a high-resolution fiber optic tactile sensing system were presented. The capabilities of the gripper and tactile sensors were validated by experimentation in underwater and deep-sea environments.

The high spatial and force resolutions of the tactile sensor system, along with the gripper and manipulator kinematics, enabled the representation of tactile data as locally dense point clouds. Thus, state-of-the-art point cloud matching techniques from range sensing literature were selected, adapted, and extended for object recognition. For adding robustness, an ICP augmented sequential methodology was developed for evolving the hypotheses generated from Batch RANSAC based database matching. The feature-based RANSAC and ICP allowed efficient management of high dimensional spaces, as well as a large object database. Free-space information, managed using a 3D occupancy map, led to an accelerated object recognition performance. The methodology is independent of the acquisition of ground truth data for database creation, and the database can be created using a simulated laser scanner. New objects can also be autonomously added to the database. Tactile data collected from underwater experiments were used to validate the algorithms for static object recognition and complete 6 DOF localization. Recognition performance of between 80% and 100% was achieved for four different objects within 20 exploration steps, for a 45-object database. The experiments showed 50–100% improvements by using the sequential framework over RANSAC and feature-based database matching. For a smaller database of five objects, a similar performance could be achieved within five grasps.

To our knowledge, haptic object recognition and 6 DOF pose estimation for structured underwater and deep-sea environments has been presented for the first time in the literature. The results presented can be considered as a step toward achieving autonomy in underwater and deep-sea environments. In future, more tactile data can be generated from pressure chamber experiments, which can then be used to test the effectiveness of the methodology for deep-sea environments. The performance is expected to be much better than the results presented assuming a worst case sensor noise of 50%. In the experiments presented, only one tactile sensor was used for object exploration. The recognition performance is expected to accelerate drastically with the use of the remaining five tactile sensors on the SeeGrip gripper. It is expected that robust recognition can be achieved within four to five grasps of such a system. This has already been tested and proven in simulation. Finally, an autonomous object exploration strategy should be used for testing the algorithms in the future. We also want to work on a closer link between the exploration strategy and the current state of recognition in the future.

ACKNOWLEDGMENTS

The results presented in this paper are based on technologies developed in the SeeGrip project, which is funded by the German Federal Ministry of Economics and Technology (BMWi) according to a resolution of the German Bundestag, Grant No. 03SX291.

REFERENCES

Aggarwal, A., & Albiez, J. (2013). Autonomous trajectory planning and following for industrial underwater manipulators. In MTS/IEEE OCEANS, San Diego.

Aggarwal, A., & Kampmann, P. (2012). Tactile sensors based object recognition and 6d pose estimation. In Intelligent Robotics and Applications (pp. 406–416). Springer.

Allen, P. (1990). Mapping haptic exploratory procedures to multiple shape representations. In Robotics and Automation, Proceedings of the IEEE International Conference (pp. 1679 –1684, vol. 3).

Allen, P. K., & Roberts, K. S. (1989). Haptic object recognition using a multi-fingered dextrous hand (pp. 342–347). New York: Department of Computer Science, Columbia University.

Arun, K. S., Huang, T. S., & Blostein, S. D. (1987). Least-squares fitting of two 3-d point sets. IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(5), 698–700.

Besl, P., & McKay, N. D. (1992). A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2), 239–256.

Besl, P. J., & Jain, R. C. (1985). Three-dimensional object recognition. ACM Computing Surveys (CSUR), 17(1), 75–145.

Bierbaum, A., Gubarev, I., & Dillmann, R. (2008a). Robust shape recovery for sparse contact location and normal data from haptic exploration. In IROS.

Bierbaum, A., Rambow, M., Asfour, T., & Dillmann, R. (2008b). A potential field approach to dexterous tactile exploration of unknown objects. In Humanoid Robots. Humanoids. 8th IEEE-RAS International Conference (pp. 360–366).

Caselli, S., Magnanini, C., & Zanichelli, F. (1994). Haptic object recognition with a dextrous hand based on volumetric shape representations. In IEEE International Conference on MFI (pp. 280–287).

Caselli, S., Magnanini, C., Zanichelli, F., & Caraffi, E. (1996). Efficient exploration and recognition of convex objects based on haptic perception. In Robotics and Automation. Proceedings of the IEEE International Conference (vol. 4, pp. 3508–3513).

Chen, N., Rink, R., & Zhang, H. (1996). Local object shape from tactile sensing. In Robotics and Automation, Proceedings of the IEEE International Conference (vol. 4, pp. 3496–3501).

Drost, B., Ulrich, M., Navab, N., & Ilic, S. (2010). Model globally, match locally: Efficient and robust 3d object recognition. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference (pp. 998–1005).

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381–395.

Gorges, N., Fritz, P., & Wörn, H. (2010a). Haptic object exploration using attention cubes. Robotics, 6359 LNAI, 349–357.

Gorges, N., Navarro, S., Göger, D., & Wörn, H. (2010b). Haptic object recognition using passive joints and haptic key features. In ICRA.

Hebert, M., Ikeuchi, K., & Delingette, H. (1995). A spherical representation for recognition of free-form surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(7), 681–690.

Hebert, P., Howard, T., Hudson, N., Ma, J., & Burdick, J. W. (2013). The next best touch for model-based localization. In Robotics and Automation (ICRA), 2013 IEEE International Conference (pp. 99–106).

Heidemann, G., & Schopfer, M. (2004). Dynamic tactile sensing for object identification. In IEEE International Conference on Robotics and Automation.

Hornung, A., Wurm, K. M., Bennewitz, M., Stachniss, C., & Burgard, W. (2013). OctoMap: An efficient probabilistic 3D mapping framework based on octrees. Autonomous Robots. Software available at http://octomap.github.com.

Hsiao, K., Kaelbling, L. P., & Lozano-Pérez, T. (2010). Task-driven tactile exploration. In Robotics: Science and systems.

Johnson, A. E., & Hebert, M. (1998). Surface matching for object recognition in complex 3-d scenes. Image and Vision Computing, 16, 635–651.

Johnsson, M., & Balkenius, C. (2007). Neural network models of haptic shape perception. Robotics and Autonomous Systems, 55(9), 720–727.

Kampmann, P., & Kirchner, F. (2012). A tactile sensing system for underwater manipulation. In Proceedings of the Work-

shop on Advances in Tactile Sensing and Touch based Human-Robot Interaction, Boston.

Kampmann, P., & Kirchner, F. (201x). Towards a fine-manipulation system with tactile feedback for deep-sea environments. Robotics and Autonomous Systems (to be published).

Kampmann, P., Lemburg, J., Hanff, H., & Kirchner, F. (2012). Hybrid pressure-tolerant electronics. In Oceans (pp. 1–5). IEEE.

Kaufman, A. E., & Shimony, S. E. (1986). 3d scan-conversion algorithms for voxel-based graphics. In Crow, F. C., & Pizer, S. M. (eds.), I3D (pp. 45–75). ACM.

Lemburg, J., Kampmann, P., & Kirchner, F. (2011). A small-scale actuator with passive-compliance for a fine-manipulation deep-sea manipulator. In OCEANS (pp. 1–4).

Mazzini, F., Kettler, D., Guerrero, J., & Dubowsky, S. (2011). Tactile robotic mapping of unknown surfaces, with application to oil wells. IEEE Transactions on Instrumentation and Measurement, 60(2), 420–429.

Meier, M., Schöpfer, M., Haschke, R., & Ritter, H. (2011). A probabilistic approach to tactile shape reconstruction. IEEE Transactions on Robotics, 27(3), 630–635.

Mian, A. S., Bennamoun, M., & Owens, R. (2005). Automatic correspondence for 3d modeling: An extensive review. International Journal of Shape Modeling, 11(2), 253.

Moll, M., & Erdmann, M. A. (2003). Reconstructing the shape and motion of unknown objects (pp. 293–310).

Muja, M., & Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. In International Conference on Computer Vision Theory and Applications (VISSAPP09) (pp. 331–340).

Park, I. K., Germann, M., Breitenstein, M. D., & Pfister, H. (2010). Fast and automatic object pose estimation for range images on the gpu. Machine Vision and Applications, 21(5), 749–766.

Paul, B. (2008). Technical concepts orientation, rotation, velocity and acceleration, and the SRM. Technical report, SEDRIS.

Petrovskaya, A., & Khatib, O. (2011). Global localization of objects via touch. IEEE Transactions on Robotics, 27(3), 569–585.

Pezzementi, Z., Reyda, C., & Hager, G. (2011). Object mapping, recognition, and localization from tactile geometry. In ICRA.

Reimer, E., & Baldwin, L. (1999). Cavity sensor technology for low cost automotive safety and control devices. In Air Bag Technology, Detroit.

Roberts, K. (1990). Robot active touch exploration: Constraints and strategies. In Robotics and Automation, Proceedings of the IEEE International Conference (pp. 980–985, vol. 2).

Rusu, R. B., Blodow, N., & Beetz, M. (2009). Fast point feature histograms (fpfh) for 3d registration. In ICRA (pp. 3212–3217). IEEE.

Rusu, R. B., & Cousins, S. (2011). 3D is here: Point cloud library (PCL). In IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China.

Rusu, R. B., Marton, Z. C., Blodow, N., & Beetz, M. (2008). Persistent point feature histograms for 3d point clouds. IAS-10 (p. 119).

Schneider, A., Sturm, J., Stachniss, C., Reisert, M., Burkhardt, H., & Burgard, W. (2009). Object identification with tactile sensors using bag-of-features. In Intelligent Robots and Systems, IEEE/RSJ International Conference (pp. 243–248).

Shan, Y., Matei, B., Sawhney, H. S., Kumar, R., Huber, D., & Hebert, M. (2004). Linear model hashing and batch ransac for rapid and accurate object recognition. In IEEE International Conference on Computer Vision and Pattern Recognition (pp. 121–128).

Shilane, P., Min, P., Kazhdan, M., & Funkhouser, T. (2004). The Princeton shape benchmark. In Shape Modeling Applications, Proceedings (pp. 167–178). IEEE.

Stein, F., & Medioni, G. (1992). Structural indexing: Efficient 3-d object recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, 125–145.

Tombari, F., Salti, S., & Di Stefano, L. (2010). Unique signatures of histograms for local surface description. In Computer Vision–ECCV (pp. 356–369). Springer.

Wahl, E., Hillenbrand, U., & Hirzinger, G. (2003). Surflet-pair-relation histograms: A statistical 3d-shape representation for rapid classification. In 3-D Digital Imaging and Modeling, 3DIM 2003. Proceedings of the Fourth International Conference (pp. 474–481). IEEE.